UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1844

A NOVEL MACHINE LEARNING APPROACH FOR IDENTIFYING THE DRIVERS OF DOMESTIC ELECTRICITY USERS' PRICE RESPONSIVENESS

Peiyang Guo Jacqueline CK Lam Victor OK Li

16 August 2018

Time-based pricing programs for domestic electricity users have been effective in reducing peak demand and facilitating renewables integration. Nevertheless, high cost, price non-responsiveness and adverse selection may create the possible challenges. To overcome these challenges, it can be fruitful to investigate the 'high-potential' users, which are more responsive to price changes and apply time-based pricing to these users. Few studies have investigated how to identify which users are more price-responsive. We aim to fill this gap by comprehensively identifying the drivers of domestic users' price responsiveness, in order to facilitate the selection of the high-potential users. We adopt a novel data-driven approach, first by a feed forward neural network model to accurately determine the baseline monthly peak consumption of individual households, followed by an integrated machine-learning variable selection methodology to identify the drivers of price responsiveness applied to Irish smart meter data from 2009-10 as part of a national Time of Use trial. This methodology substantially outperforms traditional variable selection methods by combining three advanced machine-learning techniques. Our results show that the response of energy users to price change is affected by a number of factors, ranging from demographic and dwelling characteristics, psychological factors, historical electricity consumption, to appliance ownership. In particular, historical electricity consumption, income, the number of occupants, perceived behavioural control, and adoption of specific appliances, including immersion water heater and dishwasher, are found to be significant drivers of price responsiveness. We also observe that continual price increase within a moderate range does not drive additional peak demand reduction, and that there is an intention-behaviour gap, whereby stated intention does not lead to actual peak reduction behavior. Based on our findings, we have conducted scenario analysis to demonstrate the feasibility of selecting the high potential users to achieve significant peak reduction.

A novel machine learning approach for identifying the drivers of domestic electricity users' price responsiveness

EPRG Working Paper 1824

Cambridge Working Paper in Economics 1844

Peiyang Guo, Jacqueline CK Lam, and Victor OK Li

Abstract

Time-based pricing programs for domestic electricity users have been effective in reducing peak demand and facilitating renewables integration. Nevertheless, high cost, price non-responsiveness and adverse selection may create the possible challenges. To overcome these challenges, it can be fruitful to investigate the 'high-potential' users, which are more responsive to price changes and apply time-based pricing to these users. Few studies have investigated how to identify which users are more price-responsive. We aim to fill this gap by comprehensively identifying the drivers of domestic users' price responsiveness, in order to facilitate the selection of the high-potential users. We adopt a novel data-driven approach, first by a feed forward neural network model to accurately determine the baseline monthly peak consumption of individual households, followed by an integrated machine-learning variable selection methodology to identify the drivers of price responsiveness applied to Irish smart meter data from 2009-10 as part of a national Time of Use trial. This methodology substantially outperforms traditional variable selection methods by combining three advanced machine-learning techniques. Our results show that the response of energy users to price change is affected by a number of factors, ranging from demographic and dwelling characteristics, psychological factors, historical electricity consumption, to appliance ownership. In particular, historical electricity consumption, income, the number of occupants, perceived behavioural control, and adoption of specific appliances, including immersion water heater and dishwasher, are found to be significant drivers of price responsiveness. We also observe that continual price increase within a moderate range does not drive additional peak demand reduction, and that there is an intention-behaviour gap, whereby stated intention does not lead to actual peak reduction behavior. Based on our findings, we have conducted scenario analysis to demonstrate the feasibility of selecting the high potential users to achieve significant peak reduction.

Keywords Time-based electricity pricing, price responsiveness, high-potential users, variable selection, Time of Use, machine learning

JEL Classification Q41, C55

Contact jacquelinelam@hku.hk
Publication August 2018
Financial Support Research Grants Council of Hong Kong [Grant No. RGC-GRF-17403614]

A novel machine learning approach for identifying the drivers of domestic electricity users' price responsiveness to price change

Peiyang Guo
(andrew04@hku.hk)
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Jacqueline C.K. Lam
(jcklam@eee.hku.hk)
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Victor O.K. Li
(vli@eee.hku.hk)
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

1 Introduction

Time-based electricity pricing for domestic users has been effective in reducing peak consumption, and facilitating renewable integration [1-5]. Trial studies have confirmed the effectiveness of these programs in reducing peak-time demand, with reduction ranging from approximately five percent for the simplest program of time of use (ToU), to greater than 20% for more advanced programs such as critical peak pricing and dynamic pricing [6-9]. It has also been argued that time-based pricing has the potential of improving demand flexibility and levelling out the renewable energy output variation [1, 10-12]. As such, time-based programs have proliferated in places like California, France and Northern Europe [13, 14].

Introducing time-based pricing programs, however, is not without challenges. First, most time-based programs carry with them additional costs such as hefty investments in enabling technologies, massive costs in metering and communication system upgrade, and costs in marketing and consumer enrolment [3, 4]. Second, significant price responsiveness is not observed in all households. There is a strikingly skewed distribution of price elasticity, indicating that only a fraction of the households is responsive to price change [15]. Third, time-based programs offered on an opt-in basis are liable to adverse selection, where free-rider participants provide little relief during the load-control period, but are still able to enjoy low tariffs at other intervals [4, 16-20].

One way to tackle these problems is by selectively enrolling high-potential users who are responsive to price change. However, identifying high-potential users is difficult, as there is no prior information on whether a user will be responsive. Therefore, studies have resorted to selecting several key household characteristics that are likely to make the users responsive. For example, some studies attempt to select the high-potential users by deducing the presence of certain appliances, such as heating, ventilation, and air conditioning appliances (HVAC) [21, 22], as there is evidence suggesting that users who adopt major appliances of high electricity demand are likely to be more responsive [15, 23]. Some other studies seek to select the high-potential users by segmenting the load profiles using historical consumption data [24-29]. The underlying assumption is that high-consumption users during system peak time will be more responsive to price change.

Very few research studies have attempted to identify the drivers of price responsiveness. Most peak demand studies have investigated the drivers of energy consumption and energy conservation, rather than price responsiveness [30-36]. For the small number of studies that have examined price responsiveness, they either focus on quantifying price responsiveness [7, 23, 37]; or on investigating the effects of a few factors, such as specific appliances or household income [38, 39]. Existing literature fails to provide a comprehensive account of what drives users' price responsiveness

This article aims to fill this gap by comprehensively identifying the drivers of domestic users' price responsiveness, thus facilitating the selection of high-potential users. It attempts to address the following questions. First, we want to understand which attributes of individual electricity users drive price responsiveness during the peak. Second, we want to identify what role price change will play in demand reduction. We first survey the literature on all the potential factors that might influence energy consumption behaviour. Based on these potential factors, the drivers of price responsiveness are then identified and the role of price is studied. The results can then be used to inform high-potential user selection in large-scale application.

We adopt a two-step approach to achieve the research aim. First, using a neural network model to estimate baseline energy consumption, we estimate the response to price change for each individual household, and subsequently identify the high-potential households. Second, by applying three advanced variable selection models, we identify the drivers that determine the household's responsiveness to price change. The new two-step approach provides direct and effective identification of high potential households and drivers of price responsiveness. To the best of our knowledge, as of to date, very few articles have applied machine-learning techniques to study household energy consumption given an intervention [40-42]. However, these studies have only deployed machine-learning methods to estimate "individual treatment effects" of demand response program, instead of identifying drivers that lead to price responsiveness; their model is a simple one-step model not addressing variable selection, whilst ours have deployed an integrated two-step machine-learning model that also address the challenge of variable selection.

Our approach departs from traditional methods of energy behavioural study, which have relied heavily on statistical or econometric methods to derive treatment effects from fixed effect or difference in difference modelling. For selecting significant variables, traditional methods have often relied on direct testing or step-wise selection [39, 43-45]. However, these methods have presented limitations: First, these models often have adopted linear modelling, or a modified linear modelling (for example by adding interaction terms). However, the relationship between relevant factors and household energy consumption is not always linear. For example, household energy consumption behaviours may be influenced by a complex interaction between the socio-economic variables, dwelling characteristics and appliances installed in the households. The non-linear interactions between these factors may be better captured by non-linear modelling. Second, neither direct-testing nor step-wise selection can be trusted to produce a reliable set of factors of price responsiveness. Direct testing with too many variables present in the regression can accord statistical significance to irrelevant variables, especially when some variables are correlated [46]. Stepwise selection, on the other hand, suffers from serious problems of inconsistency, as it depends largely on the algorithm used (forward or backward), the order of variable entry (or deletion) and the total number of variables[47-49].

Hence, our newly proposed integrated machine-learning model has addressed the constraints imposed by traditional models. First, our neural network model can accurately estimate the baseline peak consumption of each individual household, allowing it to learn the complex and non-linear relationship between energy consumption and other factors [50]. Second, the drivers of price responsiveness are robustly identified by combining three advanced machine-learning methods with embedded variable selection properties. These machine-learning techniques have substantial advantages over direct-testing and step-wise selection, in that they could either select relatively stable results, or solve the collinearity problem. We further use a consensus voting to select only those variables agreed by all three methods to ensure that our results of driver identification are reliable. Although we are not claiming that our methodology will overcome all the limitations of traditional methods, the combination of these methods should have greatly improved the confidence of selecting the real drivers that determine price responsiveness.

Our paper is structured as follows. Section 2 surveys the literature on potential drivers of price responsiveness to peak electricity demand. Section 3 describes the data and pre-data processing steps. Section 4 elaborates on the methodology. Section 5 presents the results. Section 6 performs the scenario analysis. Section 7 concludes our study.

2 Literature review on potential drivers, price responsiveness and peak demand

This section surveys the potential drivers of price responsiveness of domestic users.

The literature that directly investigates the drivers of price responsiveness is scattered and incomprehensive. Some early studies found the presence of energy intensive appliances to be drivers of responsiveness. In [23], the authors reviewed five early ToU experiments in the 1970s and found that households with major appliances have a significantly greater demand elasticities. A similar result was found in [15]. Some studies recorded the relevance of income and dwelling characteristics. In [15], the elasticity of the lowest income (annual income less than $US 18,000) households was found to be almost 50% higher than that of the highest income (annual income greater than $US 60,000) households. Other relevant factors recorded in the literature include weather conditions, seasonal and regional variations[23, 39, 51, 52].

Due to limited direct research on the drivers of price responsiveness, we survey instead the determinants of electricity demand. These determinants correlate closely with how users consume electricity, making them candidates for price responsiveness. Together with identified drivers from direct research, these determinants form a pool of potential drivers of price responsiveness. (Table 1).

There is much literature investigating the role of demographic and dwelling characteristics on electricity consumption. Some consensus has been reached so far regarding the role of household income, number of occupants, dwelling type, floor area and room numbers: Most studies conclude a positive relationship between electricity consumption and household income [22, 32, 53-63]; increasing the number of occupants would lead to greater electricity consumption, though it would result in lower electricity consumption per capita [32, 53, 55-59, 61, 64, 65]; electricity consumption increases with the degree of detachment of dwelling[32, 36, 53, 55, 57-59, 62, 66].

Other demographic and dwelling characteristics that have been identified from our literature survey include: family composition, age of household responsible person, social class, education level, tenure type (rent/own), dwelling age, room number[33, 36, 55, 57, 60, 63-69]. However, no consensus has been reached on the effects of these determinants. For example, [57] found that old residences would consume less electricity, possibly due to low penetration of energy-intensive appliances. In [32], however, old residences are observed to consume more electricity due to poor insulation. Some other studies [36, 65], nevertheless, reported no significant relationship between electricity consumption and dwelling age.

Weather and presence of appliances were also widely cited as the determinants of electricity consumption [33, 36, 58, 65, 70-73]. [36] ascertained that cooling degree day is the dominant determinant in the summer in total electricity consumption. In terms of appliances, dishwashers, laundry appliances (washing machine and tumble dryer), and HVAC appliances are found to have a strong statistical relationship with electricity consumption [32, 33, 58, 65, 66, 71, 72]. Interestingly, IT equipment, which nominally is not energy-intensive, is found to be taking up a sizable share of total electricity consumption. For example, results from [33] suggested that the computer was the third largest contributor to total electricity consumption in an Irish residence in 2009, second only to dishwasher and tumble dryer.

Some studies examined the role of psychological factors on electricity consumption and energy-related behaviours. In [74], perceived behavioural control is found to be the main reason for electricity curtailment behaviour such as switching off the light. In [35], energy saving is explained to a greater degree by psychological factors such as attitude and perceived control than demographic factors. In [75], social norms are also found to be positively affecting the electricity saving behaviours. These studies confirm that psychological factors have significant impact on behavioural change.

Table 1 summarizes the potential drivers of price responsiveness. Our research aims to identify those potential drivers which will make the user more responsive to price and therefore lead to more significant peak demand reduction.

Table 1 Potential drivers of price responsiveness

Categories	**Potential drivers**

Demographics	Household income, number of occupants, family composition, age of household responsible person, social class, education level, employment type, etc.
Dwelling characteristics	Dwelling type (detached house, apartment, etc.), floor area, tenure type (rent/own), house age, number of rooms, etc.
Psychological factors	Attitude, perceived control, social norms, etc.
Appliances	HVAC appliances, dishwasher, laundry appliances (washing machine and tumble dryer), computers, etc.
Weather	Heating degree days, cooling degree days, dew points, etc.
Interventions	Price, financial incentive, etc.

3 Data pre-processing

3.1 Data description

The data are derived from the smart metering trial in Ireland. The data contain the half-hour electricity consumption record from 4225 households from July 2009 to December 2010. The half year of 2009 is designated as the benchmark period, when all households are recorded for their consumption without any intervention. In 2010, households are randomly assigned into one control group and 4 treatment groups, with each treatment group receiving different ToU tariffs, as shown in Table 2 [76, 77]. The trial designated the period between 17.00 to 19.00 as the peak time, given that it was the time of the day when the total electricity demand was the highest.

Table 2 Household ToU pricing plan introduced in the smart metering trial in Ireland (cents per kWh)[1]

TIME	**BENCHMARK PERIOD (JUL. 2009- DEC. 2010)**	**TEST PERIOD (JAN. 2010-DEC. 2010)**		
		Night 23.00- 08.00	Day (08.00-17.00 and 19.00-23.00 on weekdays, 17.00-	Peak 17.00-19.00 (Monday to Friday, excluding bank holidays)

[1] There is a slight price change in October 2009 during the benchmark period because of the blanket tariff adjustment by Electric Ireland. However, the price change is meagre, only 0.2 cents/kWh. We therefore ignore this adjustment.

			19.00 on weekends and holidays)	
CONTROL			16.00	
TARIFF A	16.00	13.62	15.89	22.70
TARIFF B		12.46	15.32	29.51
TARIFF C		11.35	14.76	36.32
TARIFF D		10.22	14.19	43.13

The dataset also carries with it a comprehensive survey on household demographics, dwelling characteristics, appliances and users' attitude towards electricity curtailment.

Meteorological data in this study are retrieved from Irish National Meteorological Service.

3.2 Consumption data processing

The consumption data for each household from July 2009 to December 2010 are processed following Figure 1.

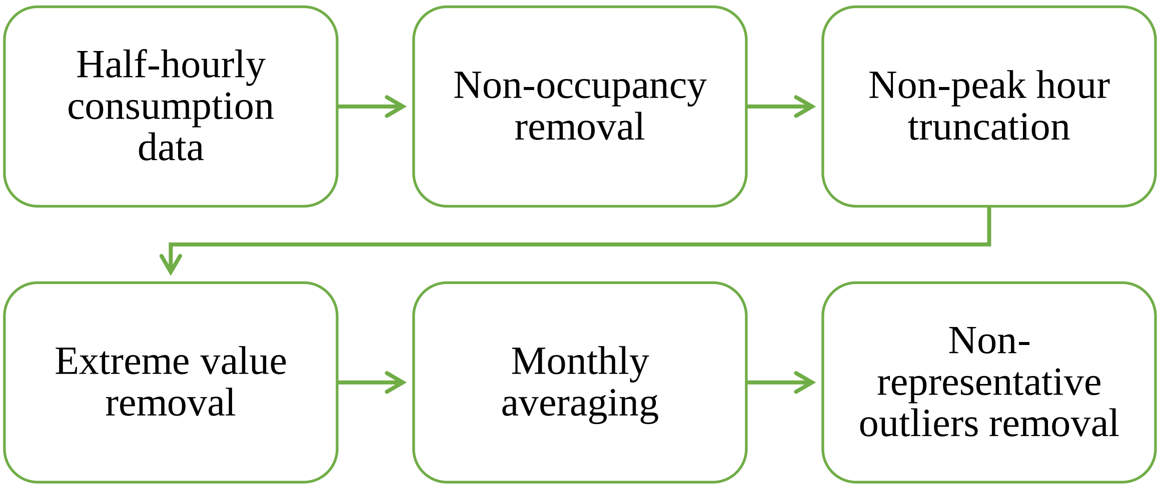

Figure 1 Flow chart for noisy electricity consumption data cleaning

1) Unoccupied day removal and non-peak hour truncation

We define a time period as non-occupied if residents are not staying at home. Only a small number of appliances are working when a residence is non-occupied (refrigerator, for example). Since our aim is to study whether and to what extent the users will change their behaviours in response to price change, the mixing of the non-occupancy data with the occupancy data would distort the users' actual behavioural response to price change. We therefore have identified the non -occupied periods and removed them from our dataset. The non-occupancy detection method follows [78] and [79], which determine the occupancy by comparing consumption features during the period of interest with the period when residents are not active (such as early hours of the morning). Refer to the Appendix for the non-occupancy detection algorithm.

Additionally, we retain only the peak period data, i.e., from 17.00 to 19.00. The reasons are as follows: 1) price variations from 2009 to 2010 are greatest during this period, and 2) demand reduction at peak time is one of the key goals of time-based pricing. Weekends and holidays are excluded because their electricity consumption profiles greatly differ from that of the weekdays [80].

2) Extreme value removal

Extreme values are peak consumptions that are extreme in magnitudes when compared to peak consumptions on other days. We use the widely used interquartile range (IQR) rule to detect the extreme values. The IQR of a data series is the difference between its first quartile (Q1) and third quartile (Q3). Extreme values are thereby defined to be values greater than $(\mathrm{Q3} + 1.5 \times \mathrm{IQR})$ or less than $(\mathrm{Q1} - 1.5 \times \mathrm{IQR})$.

3) Monthly averaging

Due to the extremely high uncertainty of energy consumption behaviour, even the most advanced deep learning cannot model long-term half-hour electricity consumption at high accuracy[81]. We therefore take the monthly average to smooth out the volatility.

4) Non-representative outlier removal

After removing non-occupied day and extreme values, some months might only contain very few unremoved days. Taking the monthly average in 3), therefore, could not even out the behavioural uncertainty. We therefore exclude months that include less days than a half of monthly working days.

3.3 Other data processing

Some key demographic factors such as income contain substantial portion of missing values because of respondents' refusal to answer. Since the effect of income has been found by many to be affecting price responsiveness (see Section 2), we therefore impute the missing income values from factors including demographics, household characteristics and appliance ownership. We adopted the Gradient Boosted Tree model and achieved a classification accuracy (the number of correct prediction/the number of total prediction) of 83% on test samples.

Psychological factors are extracted from 11 statements in the survey. Respondents indicate their degree of agreement with these statements on a 5-point Likert Scale, ranging from 'strongly agree' to 'strongly disagree'. We use factor analysis to extract the underlying factors as reflected via 11 statements. We then group together statements which have the strongest associations (loadings) with the underlying factors, and categorize each factor into relevant psychological factors that are consistent with previous psychological studies [82, 83]. A total of 4 factors are extracted: attitude indicates the respondents' overall feeling towards reducing electricity consumption; stated intention measures the stated readiness of respondents to conduct electricity reduction behaviour; perceived behavioural control describes the ease or difficulty perceived by the respondents when taking electricity consumption reduction behaviours; past behaviour assesses the efforts already invested by the respondents to reduce electricity. The reader is referred in the Appendix as the correspondence between these factors and survey statements.

4 Methodology

The objective of this study is determining, from the pool of potential drivers, which would make users more responsive to price change. The task is divided into 2 steps. The first step aims to measure the response to price change (RPC). We achieved this by building a baseline model (feed forward neural network model). The second step aims to identify which factors prompt users to have high and consistently positive RPC. We achieved this by combining three advanced machine learning techniques with variable selection properties. An overview of the methodology is shown in Figure 2.

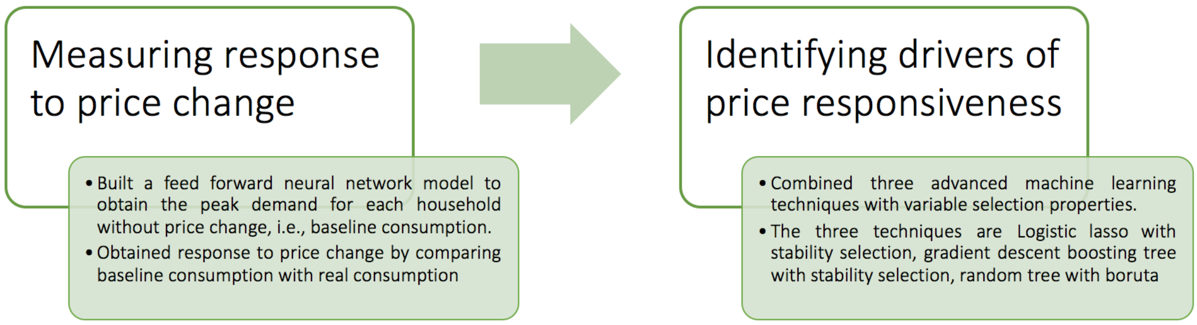

Figure 2 Methodology overview

4.1 Measuring response to price change

RPC is defined as follows:

$$RPC = \text{Consumption without price change} - \text{Consumption after price change} \quad \text{(Equation 1)}$$

While consumption after price change is recorded as the actual consumption in the treatment group, consumption without price change must be estimated. We refer to the consumption without price change as users' baseline consumption.

We built a 3-layer feed-forward neural network (FNN) as the baseline model to estimate the monthly baseline consumptions for individual households. The control group data is used as training data, because there is no price change from 2009 to 2010 within the control group. Therefore, a model trained with control group data is expected to estimate the baseline consumption in the treatment group. FNN falls under the category of supervised learning, where the model uses the training data to learn the relationship between input features and output. The model can then be applied to new input features to derive prediction. Compared with linear regression, FNN has greater ability to handle complex non-linear functions, and greater efficiency in cases where full information for the studied problem is absent[84]. So far, it has demonstrated excellent capability in predicting electricity consumption, especially electricity consumption in the long term[50, 65, 84].

The baseline consumption is derived for each individual household, rather than for a group. This differentiates our study from other work that attempts to develop a baseline, such as [85], which is constructed on a group basis.

The input and output variables are listed in Table 3. Refer to the Appendix for detailed variable definition. The input variables include all five categories of potential factors identified in Section 2. We additionally include as input six monthly consumptions in 2009. The reason for such inclusion is that historical consumption is found by many to be crucial in improving the accuracy of electricity consumption models[73, 81, 86].

Table 3 Input variables and output variables of the training step of baseline model

INPUT VARIABLE		**OUTPUT VARIABLE**
HISTORICAL CONSUMPTION	Control group: 6 monthly peak consumption from 2009/7 to 2009/12 in control group	Control group: individual monthly peak consumption from 2010/7 to 2010/12 for every household in control group.
DEMOGRAPHICS	Control group: household income, number of occupants, family composition, age of household responsible person, social-class, education level, employment type	
DWELLING CHARACTERISTIC	Control group: dwelling type, floor area, tenure type (rent/own), house age, number of rooms	
PSYCHOLOGICAL FACTORS	11 Likert Scale survey answers	
WEATHER	Heating degree days in 2010, temperature during peak hours for each month, air pressure, humidity	

APPLIANCES	Number of each appliance in 2009	

We divided the control group between the training set and the testing set, and use the R squared values to determine the accuracy of our baseline model. Numerical input data is standardized by subtracting the mean and dividing by the standard variation. Categorical input data is one-hot coded into dummy variables. For the model configuration, one hidden layer is used. Additionally, Parametric Rectified Linear Unit (PReLu) is applied as the activation function, mean squared error as the training objective and 'adam' as the optimizer. Dropout and L1 regularization are included to prevent overfitting. Figure 3 shows that the model achieved rather high accuracy; the convergence of the training and testing set indicates that the model has successfully controlled the possible overfitting problem.

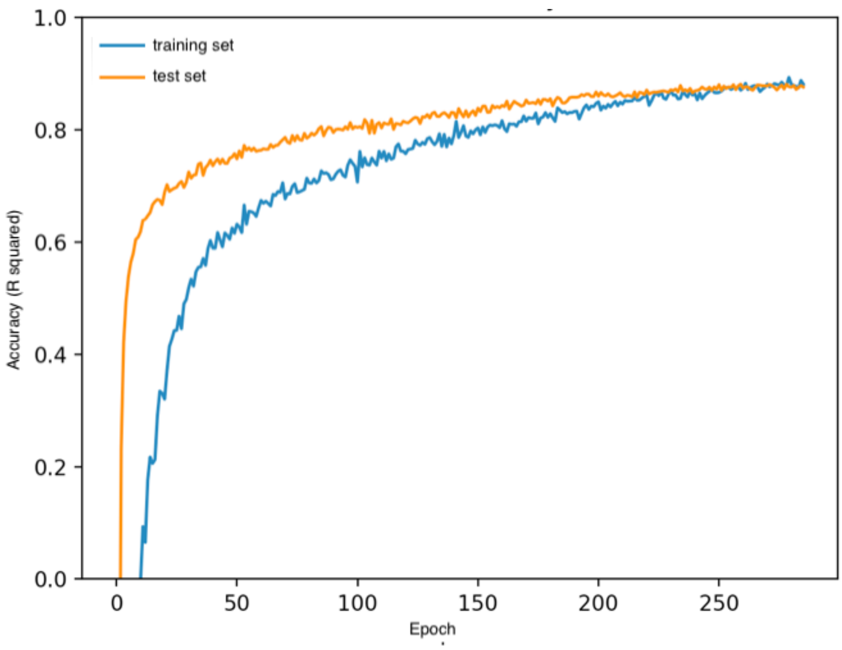

Figure 3 The accuracies of the training set and the test set of our baseline model

Overall, our baseline model has achieved an accuracy of 0.88. In comparison, if a linear regression is used on the same data, the accuracy can only reach 0.79.

After obtaining the baseline consumption for each treatment group household, we compute the RPC using Equation 1.

4.2 Identifying drivers of price responsiveness

We first define a high-potential household:

a. The average peak reduction percentage (RPC/baseline consumption) of a high-potential household is greater than the population average peak reduction percentage (reduction intensity rule).
b. The number of positive RPC of a high-potential household is greater than the average number of positive RPC in the population (reduction consistency rule).

Our aim is therefore to identify, from Table 1, factors that contribute to households becoming high-potential households. We also include historical consumption as candidate factor because it is easily obtainable by utilities and there are already researchers proposing using historical consumption to identify high-potential households[24-29]. The reader is referred to the Appendix for a detailed list of dependent and independent variables.

We approached the variable selection challenge with extreme caution and tried to select variables with consensus of multiple variable selection methods. Of the variable selection methods found in the energy consumption behavioural literature, the commonly used ones are linear regression (including mixed effects and quartile regression)[39, 60, 63, 70, 87], and stepwise regression[36, 65, 74]. These methods, however, have been found to be deeply flawed. Linear regression with too many variables will easily lead to overfitting; the multiple hypothesis testing involved in selecting variables can easily accord statistical significance to irrelevant variables[46]. Stepwise selection, on the other hand, has serious problems of parameter bias; selected variables are unstable, depending largely on the algorithm used (forward or backward), the order of variable entry (or deletion) and the total number of variables[47-49].

A new strand of variable selection methods that resort to machine learning has embedded the variable selection property. These methods, though may not be able to address all the limitations posed by traditional methods, have substantial advantages over linear regression and stepwise selection. In this part, we utilize three of these advanced machine-learning techniques and apply them to our price responsiveness study, and select factors that contribute to high-potential households by aggregating their results.

Method 1: Logistic Lasso with stability selection. Compared with stepwise selection, Lasso regression is not affected by the order of variable entry and is computationally efficient[88]. However, since Lasso regression still suffer from drawbacks such as heavy reliance on the tuning of hyper-parameters, and unstable results upon data change, we couple it with stability selection to address these problems. Simply put, stability selection bootstraps multiple times from the original datasets and performs variable selection upon each subsampling. The resulting selection probability (frequency of a variable being selected over bootstrapping iterations) is the likelihood that a variable being a true and stable variable that contributes to the prediction/determination of the response variable [89, 90]. [89] has suggested a range of (0.6, 0.9) as the cut-off probability to confirm a selected variable and that a cut-off value within this range would yield empirically similar results. Hence, we use the middle point 0.75 as the cut-off probability.

Method 2: Gradient descent boosting tree with stability selection [91]. Gradient descent boosting tree is a machine learning method that iteratively improves weak decision tree to boost performance. Compared with stepwise linear regression, gradient descent boosting tree has the advantage of modelling the non-linear relationship between response and independent variables[92]. Coupling it with stability selection enhances the capability of selecting a stable set of variables. As with method 1, we use selection probability to indicate the likelihood of a true and stable variable and choose 0.75 as the cut-off probability.

Method 3: Random Forest (RF) with Boruta. As with gradient descent boosting, RF is capable of modelling the non-linear relationships. Boruta, on the other hand, has the nice property of accommodating the collinearity of input data and finding 'all-relevant' variables, not just the 'minimal optimal' [93]. It designs a permutated 'shadow variables' for all input variables, and determines whether variables are important by subsampling and comparing between true variables and'shadow variables'. A variable is deemed to be significant (or being 'selected') if its importance score is greater than the highest importance score of 'shadow variables'. The Boruta algorithm can handle the collinearity of variables by keeping 'all-relevant' variables that are relevant to the dependent variable.

Table 4 gives an overview of the advantage of our adopted methods, and the metrics of variable significance.

Table 4 Advantages of the individual machine learning methods and their outputs.

METHOD	ADVANTAGE	OUTPUT
LOGISTIC LASSO WITH STABILITY SELECTION	Not affected by the order of variable entry; the result of variable selection is stable	A set of stable variables that contribute to the prediction/determination of the response variable.
GRADIENT DESCENT BOOSTING TREE WITH STABILITY SELECTION	Able to model the non-linear relationship; the result of variable selection is stable	A set of stable variables that contribute to the prediction/determination of the response variable.
RANDOM FOREST WITH BORUTA	Able to model the non-linear relationship; able to accommodate the collinearity of the input data	A set of confirmed variables that contribute to the prediction/determination of the response variable.

At last, we aggregated results from the three methods by consensus voting. We select only the variables that have been considered as significant in all three machine-learning methods. Our method does not presume to tackle all the challenges presented in multiple hypothesis testing and unstable results. However, each of these three methods has its own advantage over traditional methods, each can address at least part of the limitations and represents an improvement over the traditional method. By identifying the commonly agreed significant variables based on three machine-learning methods, the confidence of the selected variables as drivers of the high-potential households can increase.

5 Results and discussion

5.1 Baseline model results

Baseline model helped produce the monthly RPC of each individual household from July 2010 to December 2010. Analysing the RPCs reveal the following three results:

On the aggregate level, households achieved modest peak reduction, averaging 8.5%. This result was in line with the finding in [9] , which stated that on average ToU without user selection typically achieved a peak reduction of around 5%. This finding is also in line with the official report of the Ireland trial study, which reports an 8.33% average peak reduction[77].

While price change had produced a considerable demand reduction, there was no clear increasing trend of demand reduction when the price kept increasing. Figure 4 shows the average monthly peak demand reduction as price changes. Until now, much literature has diverged on the effect of price increasing on peak demand. While some concluded that a higher peak price/off-peak price would induce more peak demand reduction, others observe no such effect. Notably, two studies with high number of citations arrive at different conclusions. [7] reviews 15 time-based pricing experiments and finds that ‘the magnitude of price response depends on the magnitude of price increase’. However, [9] reviews 16 time-based pricing experiments in the U.S and comes to the conclusion that ‘there is no clear trend for an effect of on-peak to off-peak price ratio on peak load reduction for time of use or critical peak pricing’. Our research lends support to [9]. One explanation for this phenomenon is that when a higher peak price of moderate strength is applied, users will respond to the knowledge of such price increase. However, such response has not been rationally calculated and is therefore not strictly in proportion to the increase in price level.

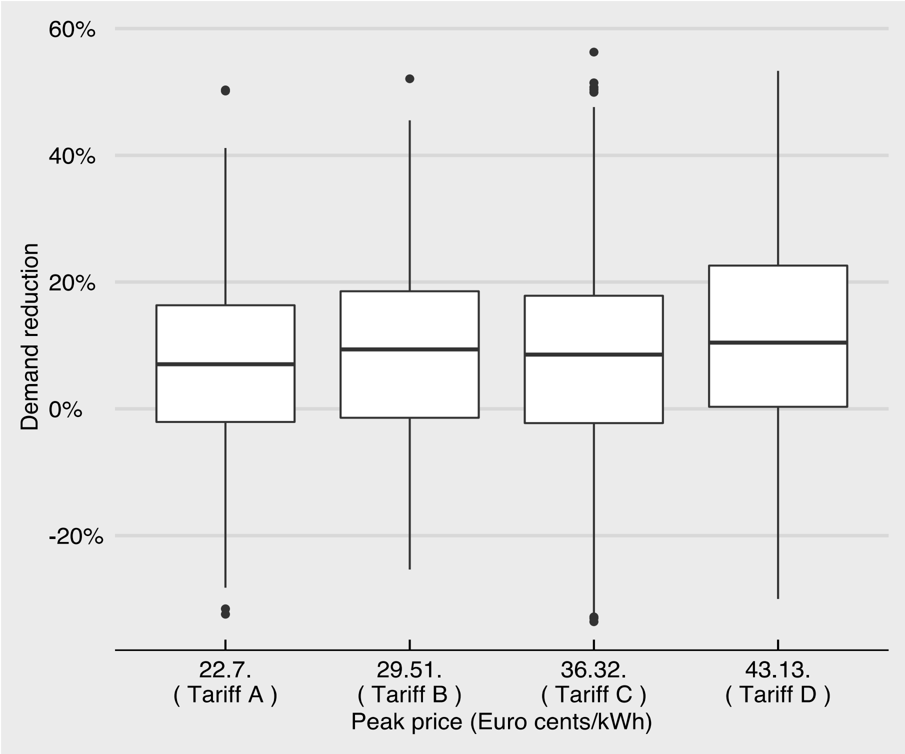

Figure 4 Effects of Different TOU Peak Prices on the Users' Peak Demand Reduction

There are considerable differences in households' response to price change. Figure 5 is a density plot that displays the distribution of average peak reduction percentage of all households. Some high-potential households achieved very high peak reduction, while some others experienced demand rise, demonstrating zero response towards price change. This again highlights the need to select the users that are responsive to price. This finding is somewhat in line with [15] , which found a highly skewed demand elasticity for energy users. The difference is that, whereas their conclusion comes from regressing *total household electricity demand* against prices under conventional tariff plans, our conclusion applies to *the peak electricity demand* in *time-based pricing*.

•

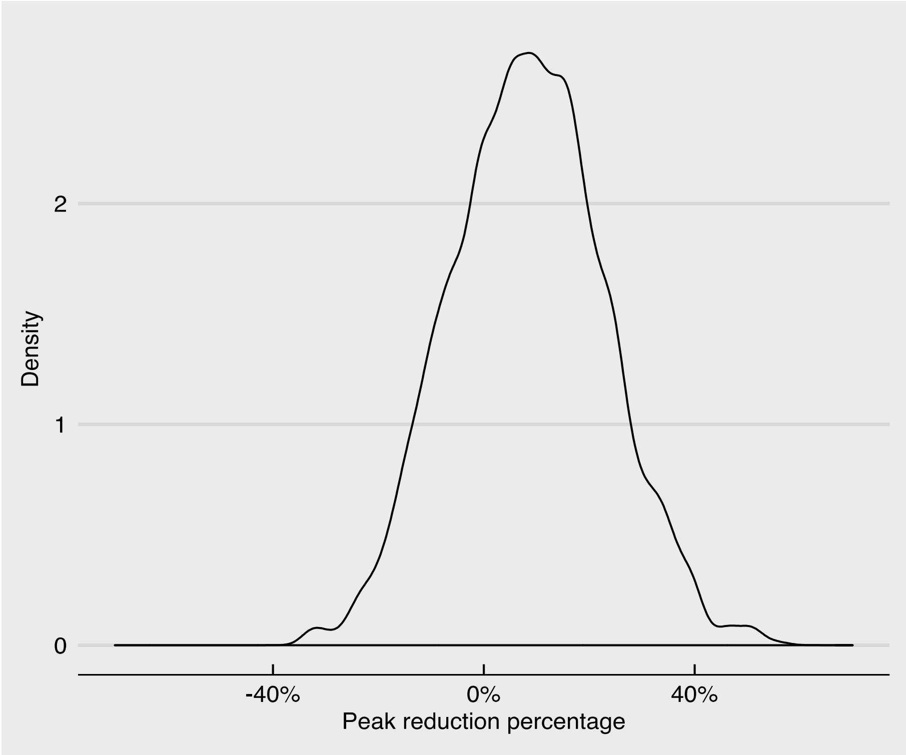

Figure 5 Distribution of the users' peak demand reduction

5.2 Results from driver identification

Figure 6 displays the variable selection results from Section 4.2. The bar represents the number of methods that have a factor contributing to households becoming high-potential households. By consensus voting, we consider factors that have a score of 3 to be of high certainty of inducing a household to respond to price change (statistically significant factors). Please refer to the Appendix for the results of individual selection methods.

It can be seen from Figure 6 that factors shown to be statistically significant encompass nearly all categories listed in Table 1. This confirms our proposition that response to price change is not driven by only one or two factors, but by a complex range of demographic and dwelling characteristics, psychological factors, and appliances.

Six factors are statistically significant, namely, historical consumption, number of occupants, income, immersion water heater, dishwasher, perceived behavioural control. A logistic regression is conducted on these 6 factors to estimate whether such factors will negatively or positively affect a household's chance of being high potential household (Table 5). We discuss these results by categories, as follows.

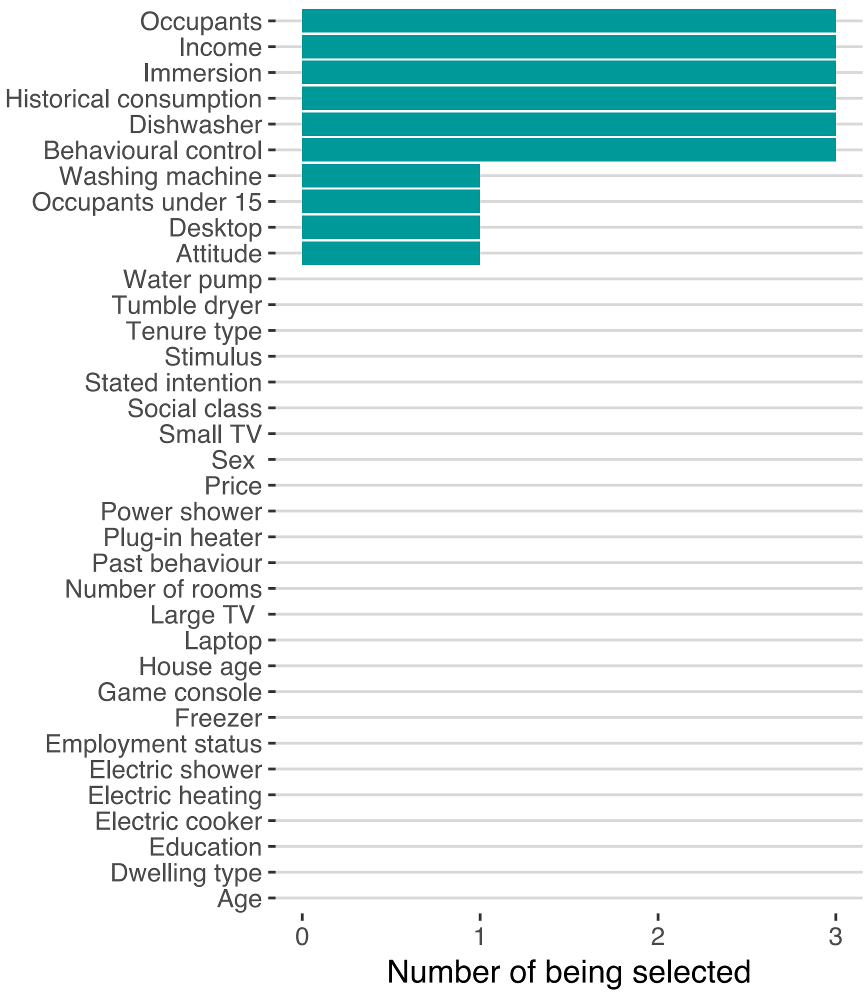

Figure 6 Drivers of peak price responsiveness of all electricity users

Table 5 Direction of the effect of the 6 most significant variables.

VARIABLE	**DIRECTION**
HISTORICAL CONSUMPTION	-
OCCUPANTS	+
INCOME	-
BEHAVIOURAL CONTROL	+

DISHWASHER	+
IMMERSION	+

1) Historic consumption

Our results show that the average level of historical consumption significantly affects the chance of households being high-potential households. It also shows that high-potential household tends to have lower average consumption. This is well expected, since we define high-potential user on percentage reduction and it is intuitive that high consumption households need more effort to achieve the same level of percentage reduction as low consumption households. This result has implication for the practice of targeting high consumption households. Utilities need to take into account the lower demand reduction percentage before deciding to enrol high consumption households.

2) Demographic and dwelling characteristics

Of all demographics and dwelling characteristics, income and number of occupants are most significant in determining whether a household is high-potential. Lower income households are more sensitive to the effect of price change to the bill and therefore are more likely to reduce demand. Higher number of occupants is also found to make a household more responsive to price change. This may be due to the fact that, when controlled for household average consumption and income, a higher number of occupants has a lower per capita income, and hence more prone to demand reduction.

3) Appliances adopted

The existence of certain appliances will improve users' control over electricity reduction behaviour and are therefore expected to affect whether a household is responsive to price change. Our results show that the number of water heating devices (immersion) and dishwasher to be significantly related to whether a household will be a high-potential household. The selection of water heating devices (immersion) and dishwasher may be explained by the fact that their usage can be flexibly postponed (dish washer) or dialled down (immersion), without incurring much inconvenience[94]. Other appliances such as freezer and washing machine do not show significance. Probable reasons may be that their load is not flexible to shift, or that shifting incurs too much inconvenience for users. Another type of appliance that conspicuously failed to show significance is space heating appliances. This may due to the fact that electric heating is seldom used as primary spacing heating method by Irish household (less than 3%), and that plug-in heaters are low in energy intensity and often used in off peak time[33].

4) Psychological factors

The perceived behavioural control is found to be strongly related to whether a user will respond to price change. The perceived behavioural control is a construct derived from two questions, one about whether the user knows what actions to take to reduce the electricity, and the other about whether the user knows which appliances should be used to reduce the electricity. It measures the perceived ease or difficulty of users to conduct demand reduction behaviours. It is therefore expected that users with high perceived behavioural control will respond with more demand reduction to price change. In comparison, stated intention is not found to be significant. This highlights the intention-behaviour gap where the intentions of people do not necessarily translate into real behaviours.

5) Price

Our results show that increasing the price do not induce a household into a responsive high-potential one. In other words, keep on increasing the price does not necessarily lead to more demand reduction. This echoes our result in Section 5.1, which observes no clear increasing trend of demand reduction when the prices keep increasing. This conclusion has great implications, as it suggests that within a moderate price range, utilities imposing a much higher price in peak time may not elicit a demand reduction significantly different from a modest price increase. However, this does not mean that users will not reduce more demand when the price is raised to an exorbitantly high level, since various trials with a critical peak pricing have successfully achieved much higher demand reduction than ToU [9].

6. Scenario analysis

Based on our results, we conduct scenario analysis on how selecting high-potential users will help reduce the peak reduction. The analysis has a simple setting, as it serves to demonstrate the feasibility and the potential of user selection, rather than accurately gauging its economic and societal benefit.

A total of two scenarios are developed: Scenario 1 is the baseline scenario in which households are enrolled randomly without selection; Scenario 2 enrols households based on our results on historical consumption, appliance ownership, demographic and dwelling characteristics. Figure 7 shows the settings of the two scenarios. Utilities can determine whether a household satisfies the criteria in Scenario 2 either by measuring and inferring from historical consumption data (for the inference of appliances, see [95-97]; for the inference of demographic and dwelling characteristics, see [98, 99]), or by consumer survey (to elicit perceived behavioural control).

To make results comparable, each scenario enrols 500 participants by resampling eligible households with replacement. The resampling is repeated 10 times and their results are averaged to ensure stability.

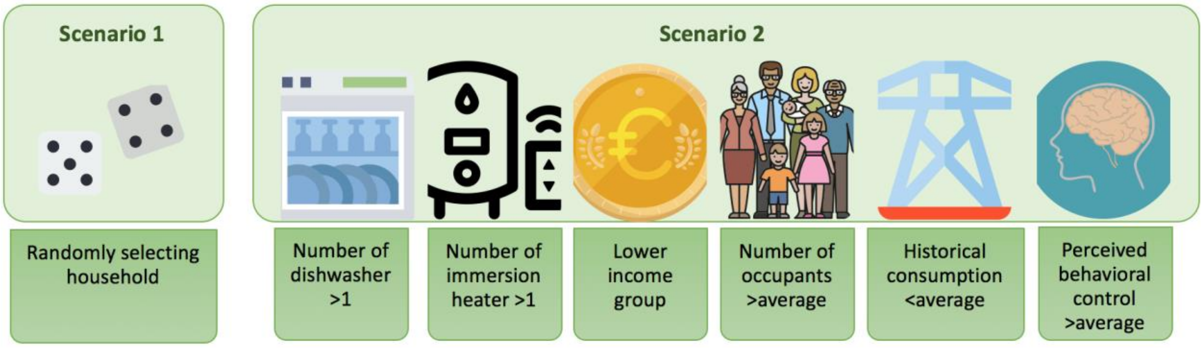

Figure 7 Scenario settings

The results are presented in Figure 8. Both the peak demand reduction and the peak reduction percentage rise considerably in Scenario 2. The peak demand reduction increases from under 10% in the baseline scenario to over 20% after user selection. This demonstrates the high potential of applying time-based pricing for selected households based on our results.

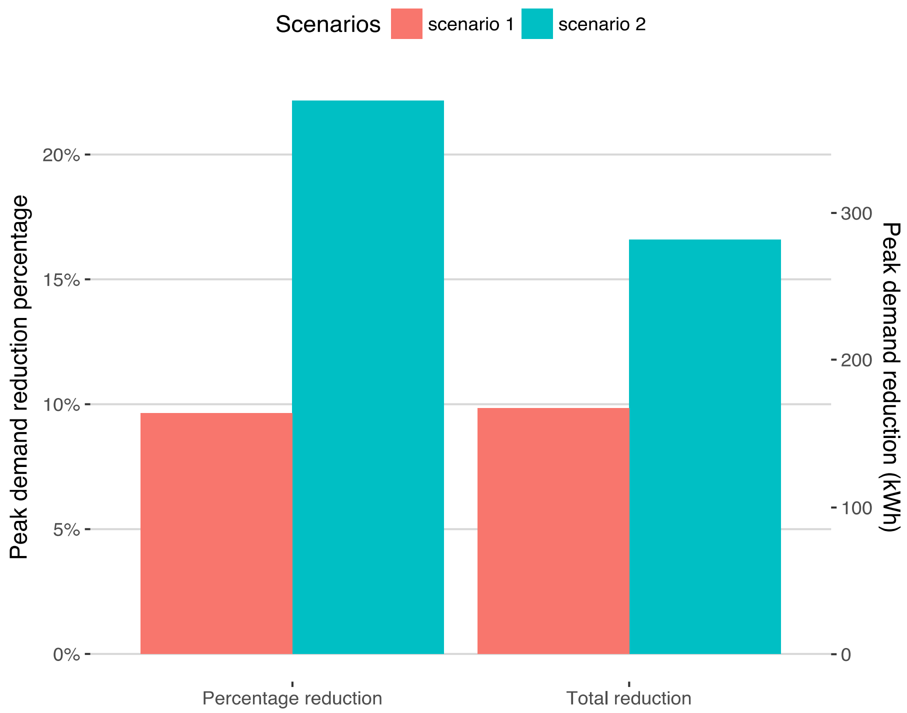

Figure 8 Results of scenario analysis

7. Conclusion

In this article, we comprehensively identified the drivers of domestic users' price responsiveness from a pool of potential factors. Our results found that price responsiveness is driven by a combination of user demographics, psychologic factors, appliance ownership and historical consumption. Our key findings include:

1) Households with higher electricity consumption usually respond less to price change
2) Ownerships of certain appliances will greatly affect the ability of users to reduce peak demand when price changes. These appliances typically are flexible to shift load to other times (dishwasher), or flexible to dial down (immersion water heater).
3) Users with better behavioural control will be more responsive to price change, while their attitude and stated intention are less significant to price responsiveness.
4) Demographic and household characteristics matter. Income and number of occupants will influence how users respond to price change.
5) Higher price increment in peak time does not necessarily elicit stronger demand reduction behaviour than lower price increment.

Our findings can provide insights on how to target the high potential users for time-based pricing. In our scenario analysis, we demonstrated the feasibility and high peak reduction of selecting users based on our results. In real world applications, targeting the high potential users and designing the corresponding time-based pricing programs is more complex, as it involves factors such as generation source optimization, balancing utilities income and user benefit, program acceptance, marketing cost, and privacy concern. Further studies are needed on appropriately introducing time-based pricing while taking these factors into consideration.

Acknowledgement

This work was supported by the Research Grants Council of Hong Kong [Grant No. RGC-GRF-17403614]. We acknowledge the provision of data from CER Smart Metering Project – Electricity Customer Behaviour Trial, 2009-2010, accessed via Irish Social Science Data Archive – www.ucd.ie/issda. We thank the invaluable comments of the anonymous reviewer.

Appendix

1. Occupancy detection

We propose an algorithm for non-occupancy detection following [78] and [79], following [78] and [79], with the following two assumptions:

- Users during the early hours of the day (02.00 to 05.00). They are either away or sleeping.
- If a residence is occupied, the consumption level and variance during the day time will be larger than that of during the night time.

Based on the two assumptions, two criteria are established:

- Average energy use during the peak hours is greater than the average energy use during inactive hours over a given period.
- Energy use variance during peak hours is greater than 1.5× average energy use variance during inactive hours over a given period[2].

We applied the algorithm for all households in the dataset and identify the period when residences are unoccupied.

2. Correspondence between psychological factors and survey statements

Factor	Statements
Attitude	It is too inconvenient to reduce our usage of electricity
	I do not have enough time to reduce my electricity usage
	I do not want to be told how much electricity I can use
	I am not able to get the people I live with to reduce their electricity usage
	Reducing my usage would not make enough of a difference to my bill
Stated intention	I/We can reduce my electricity bill by changing the way the people I/we live with use electricity
	I/We would like to do more to reduce electricity usage
Perceived behavioural control	I/We know what I/we need to do in order to reduce electricity usage

[2] This 1.5 × criterion is established by trial and error. This follows the precedence work of [280] which tried to give a fixed variance level by observing the accuracy of occupancy detection.

	I do not know enough about how much electricity different appliances use in order to reduce my usage
Past behaviour	I/We have already done a lot to reduce the amount of electricity I/we use
	I/We have already made changes to the way II/we live my life in order to reduce the amount of electricity we use.

3. Input and output variables of the baseline model

Output variables		**Definition**
	Consumption	Monthly peak consumption from 2010/7 to 2010/12 for every household in the control group.

Input categories	**Input variables**	**Definition**
Historical consumption	Historical consumption	Historical monthly peak time consumption from 2009/7 to 2009/12 in the control group
Demographics	Occupants	Total number of occupants in the household.
	Occupants under 15	Number of occupants in the household under the age of 15.
	Education	Education of chief income owner , with 1- No formal education; 2- Primary; 3- Secondary to intermediate Cert Junior Cert level; 4 -Secondary to Leaving Cert level; 5- Third Tertiary level.
	Income	Household income, with 1- Under 30,000 Euro/year; 2-Above 30,000 Euro/year.
	Sex	Sex of the respondent, with 1-Male; 2-Female.
	Age	Age of the respondent, 1- (18-25); 2- (26-35); 3- (36-45); 4- (46-55); 5- (56-65); 6-65+.

	Employment status	Employment status of chief income owner, with 1- An employee; 2- Self-employed with employees; 3- Self-employed with no employees; 4- Unemployed (actively seeking work); 5- Unemployed (not actively seeking work); 6-Retired; 7- Carer (looking after relative family).
	Social class	Social class of chief income owner, with 1-AB; 2-C1; 3-C2; 4-DE; 5-Farmers. The grade of A, B, C1, C2, D and E are defined per National Readership Survey Social Grade).
Dwelling characteristics	Dwelling type	Type of the dwelling, with 1- Apartment; 2-Semi-detached house; 3- Detached house; 4- Terraced house; 5-Bungalow.
	Tenure type	Type of the dwelling tenure, with 1- Rent; 2-Owned.
	Number of rooms	Number of bedrooms.
	House age	Age of the dwelling, with 1- Less than 5 years old; 2- Less than 10 years old; 3- Less than 30 years old; 4- Less than 75 years old; 5- Over 75 years old
Psychological factors	Indicated agreement (total number of 11)	Indicated degree of agreement with the 11 statements in the survey. Refer to section 2 of this appendix for the 11 statements.
Appliances	Electric heating	Presence of electric central heating system, with 0- No; 1-Yes.
	Washing machine	Number of washing machines
	Tumble dryer	Number of tumble dryers
	Dishwasher	Number of dishwashers
	Electric shower	Number of electric showers (instant)
	Power shower	Number of electric showers (electric hot water pumped from hot water tank)

	Electric cooker	Number of electric cookers
	Plug-in heater	Number of plug-in space heaters
	Freezer	Number of freezers
	Water pump	Number of water pumps or electric well pumps or pressurised water systems
	Immersion	Number of immersion water heaters
	Small TV	Number of TVs less than 21 inch
	Large TV	Number of TV greater than 21 inch
	Desktop	Number of desktop computers
	Laptop	Number of laptop computers
	Game console	Number of game consoles
Weather	Degree days	Monthly heating degree days in 2010
	Precipitation	Monthly average precipitation during occupied peak hour in 2010
	Temperature	Monthly average temperature during occupied peak hour in 2010
	Humidity	Monthly average humidity during occupied peak hour in 2010

4. Dependent and independent variables of the variable selection

Dependent variables		**Definition**
Household peak reduction potential		Whether the user is a high potential user, with 0-No; 1-Yes.
Independent categories	**Independent variables**	**Definition**

Historical consumption	Historical consumption	Historical monthly peak time consumption from 2009/7 to 2009/12 of treatment group households
Demographics	Occupants	Total number of occupants in the household.
	Occupants under 15	Number of occupants in the household under the age of 15.
	Education	Education of chief income owner , with 1- No formal education; 2- Primary; 3- Secondary to intermediate Cert Junior Cert level; 4 -Secondary to Leaving Cert level; 5- Tertiary (third) level.
	Income	Household income, with 1- Under 30,000 Euro/year; 2-Above 30,000 Euro/year.
	Sex	Sex of the respondent, with 1-Male; 2-Female.
	Age	Age of the respondent, 1- (18-25); 2- (26-35); 3- (36-45); 4- (46-55); 5- (56-65); 6- 65+.
	Employment status	Employment status of chief income owner, with 1- An employee; 2- Self-employed with employees; 3- Self-employed with no employees; 4- Unemployed (actively seeking work); 5- Unemployed (not actively seeking work); 6-Retired; 7- Carer (looking after relative family).
	Social class	Social class of chief income owner, with 1- AB; 2- C1; 3-C2; 4-DE; 5-Farmers. The grade of A, B, C1, C2, D and E are defined per National Readership Survey Social Grade).
Dwelling characteristics	Dwelling type	Type of the dwelling, with 1- Apartment; 2- Semi-detached house; 3- Detached house; 4- Terraced house; 5-Bungalow.
	Tenure type	Type of the dwelling tenure, with 1- Rent; 2- Owned.
	Number of rooms	Number of bedrooms.

	House age	Age of the dwelling, with 1- Less than 5 years old; 2- Less than 10 years old; 3- Less than 30 years old; 4- Less than 75 years old; 5- Over 75 years old
Psychological factors	Attitude	Respondents' overall feeling toward reducing electricity consumption
	Stated intention	Stated readiness of respondents to conduct electricity reduction behaviour
	Behavioural control	Perceived behavioural control. The ease or difficulty perceived by respondents when taking electricity consumption reduction behaviours
	Past behaviour	The effort already invested by the respondent to reduce electricity
Appliances	Electric heating	Presence of electric central heating system, with 0- No; 1-Yes.
	Washing machine	Number of washing machines
	Tumble dryer	Number of tumble dryers
	Dishwasher	Number of dishwashers
	Electric shower	Number of electric showers (instant)
	Power shower	Number of electric showers (electric hot water pumped from hot water tank)
	Electric cooker	Number of electric cookers
	Plug-in heater	Number of plug-in space heaters
	Freezer	Number of freezers
	Water pump	Number of water pumps or electric well pumps or pressurised water systems
	Immersion	Number of immersion water heaters
	Small TV	Number of TVs less than 21 inch
	Large TV	Number of TV greater than 21 inch

Desktop	Number of desktop computers
Laptop	Number of laptop computers
Game console	Number of game consoles

5. Results of 3 variable selection methods

Variables	**Method1**	**Method 2**	**Method 3**
Historical consumption	1	1	1
Occupants	1	1	1
Occupants under 15	0	0	1
Education	0	0	0
Income	1	1	1
Sex	0	0	0
Age	0	0	0
Employment status	0	0	0
Social class	0	0	0
Dwelling type	0	0	0
Tenure type	0	0	0
Number of rooms	0	0	0
House age	0	0	0
Attitude	0	0	1
Stated intention	0	0	0

Behavioural control	1	1	1
Past behaviour	0	0	0
Electric heating	0	0	0
Washing machine	0	0	1
Tumble dryer	0	0	0
Dishwasher	1	1	1
Electric shower	0	0	0
Power shower	0	0	0
Electric cooker	0	0	0
Plug-in heater	0	0	0
Freezer	0	0	0
Water pump	0	0	0
Immersion	1	1	1
Small TV	0	0	0
Large TV	0	0	0
Desktop	0	1	0
Laptop	0	0	0

Note: 1 represents being selected; 0 represent not selected.

References:

[1] Aghaei J, Alizadeh M-I. Demand response in smart electricity grids equipped with renewable energy sources: A review. Renewable and Sustainable Energy Reviews. 2013;18:64-72.

[2] Balijepalli VM, Pradhan V, Khaparde S, Shereef R. Review of demand response under smart grid paradigm. Innovative Smart Grid Technologies-India (ISGT India), 2011 IEEE PES: IEEE; 2011. p. 236-43.
[3] Bradley P, Leach M, Torriti J. A review of the costs and benefits of demand response for electricity in the UK. Energy Policy. 2013;52:312-27.
[4] Guo P, Li VO, Lam JC. Smart demand response in China: Challenges and drivers. Energy Policy. 2017;107:1-10.
[5] Pinson P, Madsen H. Benefits and challenges of electrical demand response: A critical review. Renewable and Sustainable Energy Reviews. 2014;39:686-99.
[6] Cappers P, Goldman C, Kathan D. Demand response in US electricity markets: Empirical evidence. Energy. 2010;35:1526-35.
[7] Faruqui A, Sergici S. Household response to dynamic pricing of electricity: a survey of 15 experiments. Journal of regulatory Economics. 2010;38:193-225.
[8] Stromback J, Dromacque C, Yassin MH, VaasaETT GETT. The potential of smart meter enabled programs to increase energy and systems efficiency: a mass pilot comparison Short name: Empower Demand. Vaasa ETT. 2011.
[9] Newsham GR, Bowker BG. The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: a review. Energy policy. 2010;38:3289-96.
[10] Cappers P, Mills A, Goldman C, Wiser R, Eto JH. An assessment of the role mass market demand response could play in contributing to the management of variable generation integration issues. Energy Policy. 2012;48:420-9.
[11] De Jonghe C, Hobbs BF, Belmans R. Value of price responsive load for wind integration in unit commitment. IEEE Transactions on power systems. 2014;29:675-85.
[12] Sioshansi R, Short W. Evaluating the impacts of real-time pricing on the usage of wind generation. IEEE Transactions on Power Systems. 2009;24:516-24.
[13] COWI. Impact assessment study on downstream flexibility, price flexibility, demand response & smart metering. Brussels: European Commission-DG Energy; 2016.
[14] FERC. Assessment of demand response and advanced metering: Staff report. Federal Energy Regulatory Commission; 2016.
[15] Reiss PC, White MW. Household electricity demand, revisited. The Review of Economic Studies. 2005;72:853-83.
[16] Herter K. Residential implementation of critical-peak pricing of electricity. Energy Policy. 2007;35:2121-30.

[17] Woo C-K, Kollman E, Orans R, Price S, Horii B. Now that California has AMI, what can the state do with it? Energy Policy. 2008;36:1366-74.
[18] Horowitz I, Woo C-K. Designing Pareto-superior demand-response rate options. Energy. 2006;31:1040-51.
[19] Hartway R, Price S, Woo C-K. Smart meter, customer choice and profitable time-of-use rate option. Energy. 1999;24:895-903.
[20] Woo C-K, Horii B, Horowitz I. The Hopkinson tariff alternative to TOU rates in the Israel Electric Corporation. Managerial and Decision Economics. 2002;23:9-19.
[21] Albert A, Rajagopal R. Finding the right consumers for thermal demand-response: An experimental evaluation. IEEE Transactions on Smart Grid. 2016.
[22] Kouzelis K, Tan ZH, Bak-Jensen B, Pillai JR, Ritchie E. Estimation of residential heat pump consumption for flexibility market applications. IEEE Transactions on Smart Grid. 2015;6:1852-64.
[23] Caves DW, Christensen LR, Herriges JA. Consistency of residential customer response in time-of-use electricity pricing experiments. Journal of Econometrics. 1984;26:179-203.
[24] Wang Y, Chen Q, Kang C, Zhang M, Wang K, Zhao Y. Load profiling and its application to demand response: A review. Tsinghua Science and Technology. 2015;20:117-29.
[25] McLoughlin F, Duffy A, Conlon M. A clustering approach to domestic electricity load profile characterisation using smart metering data. Applied energy. 2015;141:190-9.
[26] Rhodes JD, Cole WJ, Upshaw CR, Edgar TF, Webber ME. Clustering analysis of residential electricity demand profiles. Applied Energy. 2014;135:461-71.
[27] Al-Wakeel A, Wu J, Jenkins N. K-means based load estimation of domestic smart meter measurements. Applied Energy. 2016.
[28] Wang Y, Chen Q, Kang C, Xia Q. Clustering of electricity consumption behavior dynamics toward big data applications. IEEE Transactions on Smart Grid. 2016;7:2437-47.
[29] Al-Otaibi R, Jin N, Wilcox T, Flach P. Feature Construction and Calibration for Clustering Daily Load Curves from Smart-Meter Data. IEEE Transactions on Industrial Informatics. 2016;12:645-54.
[30] Karlin B, Davis N, Sanguinetti A, Gamble K, Kirkby D, Stokols D. Dimensions of conservation: exploring differences among energy behaviors. Environment and Behavior. 2014;46:423-52.
[31] Urban J, Ščasný M. Exploring domestic energy-saving: The role of environmental concern and background variables. Energy policy. 2012;47:69-80.

[32] Leahy E, Lyons S. Energy use and appliance ownership in Ireland. Energy Policy. 2010;38:4265-79.
[33] McLoughlin F, Duffy A, Conlon M. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. Energy and Buildings. 2012;48:240-8.
[34] O'Doherty J, Lyons S, Tol RS. Energy-using appliances and energy-saving features: Determinants of ownership in Ireland. Applied Energy. 2008;85:650-62.
[35] Abrahamse W, Steg L. How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? Journal of economic psychology. 2009;30:711-20.
[36] Kavousian A, Rajagopal R, Fischer M. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. Energy. 2013;55:184-94.
[37] Wolak FA. An experimental comparison of critical peak and hourly pricing: the PowerCentsDC program. Department of Economics Stanford University. 2010.
[38] Gyamfi S, Krumdieck S, Urmee T. Residential peak electricity demand response—Highlights of some behavioural issues. Renewable and Sustainable Energy Reviews. 2013;25:71-7.
[39] Thorsnes P, Williams J, Lawson R. Consumer responses to time varying prices for electricity. Energy Policy. 2012;49:552-61.
[40] Balandat M. New tools for econometric analysis of high-frequency time series data-application to demand-side management in electricity markets: UC Berkeley; 2016.
[41] Zhou D, Balandat M, Tomlin C. Residential demand response targeting using machine learning with observational data. Decision and Control (CDC), 2016 IEEE 55th Conference on: IEEE; 2016. p. 6663-8.
[42] Zhou DP, Balandat M, Tomlin CJ. Estimating Treatment Effects of a Residential Demand Response Program Using Non-experimental Data. Data Mining Workshops (ICDMW), 2017 IEEE International Conference on: IEEE; 2017. p. 95-102.
[43] Fujimi T, Kajitani Y, Chang SE. Effective and persistent changes in household energy-saving behaviors: Evidence from post-tsunami Japan. Applied Energy. 2016;167:93-106.
[44] Iwafune Y, Yagita Y. High-resolution determinant analysis of Japanese residential electricity consumption using home energy management system data. Energy and Buildings. 2016;116:274-84.

[45] Stewart J, Todd A. Residential behavior evaluation protocol. National Renewable Energy Laboratory; 2017.
[46] Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. Deutsches Ärzteblatt International. 2010;107:776.
[47] Ratner B. Variable selection methods in regression: Ignorable problem, outing notable solution. Journal of Targeting, Measurement and Analysis for Marketing. 2010;18:65-75.
[48] Olusegun AM, Dikko HG, Gulumbe SU. Identifying the limitation of stepwise selection for variable selection in regression analysis. American Journal of Theoretical and Applied Statistics. 2015;4:414-9.
[49] Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP. Why do we still use stepwise modelling in ecology and behaviour? Journal of animal ecology. 2006;75:1182-9.
[50] Suganthi L, Samuel AA. Energy models for demand forecasting—A review. Renewable and sustainable energy reviews. 2012;16:1223-40.
[51] Bernstein MA, Griffin J. Regional differences in the price-elasticity of demand for energy. National Renewable Energy Laboratory (NREL), Golden, CO.; 2006.
[52] Aigner D. The residential electricity time-of-use pricing experiments: What have we learned? Social experimentation: University of Chicago Press; 1985. p. 11-54.
[53] Yohanis YG. Domestic energy use and householders' energy behaviour. Energy Policy. 2012;41:654-65.
[54] Moll HC, Noorman KJ, Kok R, Engström R, Throne‐Holst H, Clark C. Pursuing more sustainable consumption by analyzing household metabolism in European countries and cities. Journal of Industrial Ecology. 2005;9:259-75.
[55] Gram-Hanssen K, Kofod C, Petersen KN. Different everyday lives: different patterns of electricity use. 2004 ACEEE Summer study on energy efficiency in buildings2004. p. 1-13.
[56] Zhou S, Teng F. Estimation of urban residential electricity demand in China using household survey data. Energy Policy. 2013;61:394-402.
[57] Tiwari P. Architectural, demographic, and economic causes of electricity consumption in Bombay. Journal of Policy Modeling. 2000;22:81-98.
[58] Bartiaux F, Gram-Hanssen K. Socio-political factors influencing household electricity consumption: A comparison between Denmark and Belgium. ECEEE Summer Study Proceedings2005. p. 1313-25.
[59] Wiesmann D, Azevedo IL, Ferrão P, Fernández JE. Residential electricity consumption in Portugal: Findings from top-down and bottom-up models. Energy Policy. 2011;39:2772-9.

[60] Sanquist TF, Orr H, Shui B, Bittner AC. Lifestyle factors in US residential electricity consumption. Energy Policy. 2012;42:354-64.
[61] Bedir M, Hasselaar E, Itard L. Determinants of electricity consumption in Dutch dwellings. Energy and buildings. 2013;58:194-207.
[62] Pachauri S. An analysis of cross-sectional variations in total household energy requirements in India using micro survey data. Energy policy. 2004;32:1723-35.
[63] Huang W-H. The determinants of household electricity consumption in Taiwan: Evidence from quantile regression. Energy. 2015;87:120-33.
[64] Ndiaye D, Gabriel K. Principal component analysis of the electricity consumption in residential dwellings. Energy and buildings. 2011;43:446-53.
[65] Tso GK, Yau KK. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy. 2007;32:1761-8.
[66] Jones RV, Fuertes A, Lomas KJ. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. Renewable and Sustainable Energy Reviews. 2015;43:901-17.
[67] Hamilton IG, Steadman PJ, Bruhns H, Summerfield AJ, Lowe R. Energy efficiency in the British housing stock: energy demand and the homes energy efficiency database. Energy Policy. 2013;60:462-80.
[68] Wyatt P. A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England. Energy Policy. 2013;60:540-9.
[69] Druckman A, Jackson T. Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model. Energy Policy. 2008;36:3177-92.
[70] Baker KJ, Rylatt RM. Improving the prediction of UK domestic energy-demand using annual consumption-data. Applied Energy. 2008;85:475-82.
[71] Larsen BM, Nesbakken R. Household electricity end-use consumption: results from econometric and engineering models. Energy Economics. 2004;26:179-200.
[72] Carter A, Craigwell R, Moore W. Price reform and household demand for electricity. Journal of Policy Modeling. 2012;34:242-52.
[73] Kim T, Lee D, Choi J, Spurlock A, Sim A, Todd A, et al. Extracting Baseline Electricity Usage Using Gradient Tree Boosting. 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). Chengdu2015.
[74] Botetzagias I, Malesios C, Poulou D. Electricity curtailment behaviors in Greek households: Different behaviors, different predictors. Energy Policy. 2014;69:415-24.

[75] Wang Z, Zhang B, Yin J, Zhang Y. Determinants and policy implications for household electricity-saving behaviour: evidence from Beijing, China. Energy Policy. 2011;39:3550-7.
[76] Di Cosmo V, Lyons S, Nolan A. Estimating the impact of time-of-use pricing on Irish electricity demand. 2012.
[77] CER. Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report. Dublin: Commission for Energy Regulation; 2011.
[78] Chen D, Barker S, Subbaswamy A, Irwin D, Shenoy P. Non-intrusive occupancy monitoring using smart meters. Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings: ACM; 2013. p. 1-8.
[79] Williams S, Gask K. Modelling sample data from smart-type electricity meters to assess potential within official statistics. Newport, South Wales, UK: The office for National Statistics, UK; 2015.
[80] Anderson B, Lin S, Newing A, Bahaj A, James P. Electricity consumption and household characteristics: Implications for census-taking in a smart metered future. Computers, Environment and Urban Systems. 2017;63:58-67.
[81] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network. IEEE Transactions on Smart Grid. 2017.
[82] Sommer L. The theory of planned behaviour and the impact of past behaviour. The International Business & Economics Research Journal. 2011;10:91.
[83] Ajzen I. The theory of planned behavior. Organizational behavior and human decision processes. 1991;50:179-211.
[84] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks. Energy. 2010;35:512-7.
[85] Schofield J, Garmichael R, Tindemans S, Woolf M, Bilton M, Strbac G. Residential consumer responsiveness to time-varying pricing. London: LCNF project: Imperial College London; 2014.
[86] Ghofrani M, Hassanzadeh M, Etezadi-Amoli M, Fadali M. Smart meter based short-term load forecasting for residential customers. North American Power Symposium (NAPS), 2011: IEEE; 2011. p. 1-5.
[87] Brounen D, Kok N, Quigley JM. Residential energy use and conservation: Economics and demographics. European Economic Review. 2012;56:931-45.
[88] Hesterberg T, Choi NH, Meier L, Fraley C. Least angle and $\ell 1$ penalized regression: A review. Statistics Surveys. 2008;2:61-93.

[89] Meinshausen N, Bühlmann P. Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2010;72:417-73.
[90] Shah RD, Samworth RJ. Variable selection with error control: another look at stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2013;75:55-80.
[91] Hofner B, Boccuto L, Göker M. Controlling false discoveries in high-dimensional situations: Boosting with stability selection. BMC bioinformatics. 2015;16:144.
[92] Mayr A, Binder H, Gefeller O, Schmid M. The evolution of boosting algorithms. Methods of information in medicine. 2014;53:419-27.
[93] Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36:1-13.
[94] Seebach D, Timpe C, Bauknecht D. Costs and benefits of smart appliances in Europe. Öko-Institut eV. 2009.
[95] Barker S, Musthag M, Irwin D, Shenoy P. Non-intrusive load identification for smart outlets. Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on: IEEE; 2014. p. 548-53.
[96] Reinhardt A, Burkhardt D, Zaheer M, Steinmetz R. Electric appliance classification based on distributed high resolution current sensing. Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference on: IEEE; 2012. p. 999-1005.
[97] Barsim KS, Mauch L, Yang B. Neural Network Ensembles to Real-time Identification of Plug-level Appliance Measurements. arXiv preprint arXiv:180206963. 2018.
[98] Gajowniczek K, Ząbkowski T. Data mining techniques for detecting household characteristics based on smart meter data. Energies. 2015;8:7407-27.
[99] Beckel C, Sadamori L, Staake T, Santini S. Revealing household characteristics from smart meter data. Energy. 2014;78:397-410.